Comparison and Evaluation of Sonification Strategies for Guidance Tasks

Gaetan Parseihian, Charles Gondre, Mitsuko Aramaki, Sølvi Ystad, Richard Kronland-Martinet

Comparison and evaluation of sonification strategies for guidance tasks

Gaëtan Parseihian, Charles Gondre, Mitsuko Aramaki, *Senior Member, IEEE,* Sølvi Ystad, Richard Kronland Martinet, *Senior Member, IEEE*

Abstract

This article aims to reveal the efficiency of sonification strategies in terms of rapidity, precision and overshooting in the case of a one-dimensional guidance task. The sonification strategies are based on the four main perceptual attributes of a sound (i.e. pitch, loudness, duration/tempo and timbre) and classified with respect to the presence or not of one or several auditory references. Perceptual evaluations are used to display the strategies in a precision/rapidity space and enable prediction of user behavior for a chosen sonification strategy. The evaluation of sonification strategies constitutes a first step toward general guidelines for sound design in interactive multimedia systems that involve guidance issues.

Index Terms

Sonification, guidance, auditory feedback, parameter mapping sonification.

I. Introduction

Driving a car or riding a bike blind, directly finding one's way in an unfamiliar smoky environment or being able to improve the quality of one's gestures in real-time ... Such ideas sound utopian, unfeasible, or based on science fiction. Meanwhile, a growing number of applications involving the auditory modality to inform or guide users can be found in a large number of domains ranging from pedestrian navigation aid [1], [2], or hand guidance for visually impaired [3], to positional guidance in surgery [4], [5], rehabilitation for patients with disabilities in medicine [6], [7], eyes free navigation in graphical user interfaces [8] or mobile devices [9], or guidance to increase the efficiency of athletes' movements [10]. In spite of the increasing interest for auditory guidance, fundamental studies on the efficiency of specific sound attributes

All authors are with the LMA, CNRS, UPR 7051, Aix-Marseille Univ, Centrale Marseille, France e-mail: parseihian@lma.cnrs-mrs.fr.

to guide a user towards a target seems to be missing. Hence, this study aims at evaluating the efficiency of such attributes in order to propose new guidelines for future applications involving auditory stimuli.

Defined in [11], [12] as "the use of non-speech audio to convey information or perceptual data", sonification constitutes a relevant method to approach these guidance issues. Indeed, it exploits the strong ability of the auditory system to analyze dynamic information, to recognize temporal or frequency changes and patterns, and to process multiple auditory streams at the same time [13], [14], [15]. Sonification is also used to improve information transmission when human modalities are not sensible to the data (i.e. reveal invisible phenomena such as radioactivity) or when the cognitive load should be limited in other modalities (i.e. adding non-visual information to a driver). It is thus an ideal candidate for guidance applications. In addition, due to the strong relationship that exists between the auditory and sensorimotor systems [16], [17], it is also a valuable tool for applications that involve perception of one's own body motion. Using sounds to guide the user in a specific task involves interaction with continuous sounds in control loops. In such cases, user actions are continuously transformed into sonic feedback. This process is called interactive sonification [18], [19] and has been shown to be very effective, for example, in enhancing the performance of the human perceptual system in the field of motor control and motor learning [20] or in enhancing 3D navigation in virtual environments [21].

Depending on the application, the audio feedback should guide the user on one or several variables (e.g. speed and steering wheel angle in a driving aid application) and/or in one or several spatial dimensions (1D, 2D, 3D in Cartesian or spherical coordinates). Depending on the context, guidance may also be either directed toward a static (e.g. guiding the user's hand to grasp an object) or a dynamic target (e.g. a pursuit-tracking task). All these considerations strongly influence the choice of the mapping between the data and the sound.

Different methods for converting guidance information into audible streams have been used since the creation of the sonification research field. Globally these methods can be classified in two paradigms based on association between data and sounds. The first paradigm ("spatial sonification"), is based on human perceptual and cognitive capacities for spatial hearing to guide the user. It consists in using the natural abilities of the hearing system to locate the target position by virtually rendering its position through stereophony or virtual auditory display (such as 3D binaural rendering) [22]. Most electronic travel aids for visually impaired are based on this paradigm and display guidance information with spatialized auditory icons or earcons [1], [2], [23]. The second category ("non-spatial sonification"), uses perceptual characteristics of sounds, such as pitch, loudness, tempo, brightness, fluctuation strength, etc. to transmit guidance information to the user. In this category, the link between the information and the

auditory display is metaphorical (there is no natural connection between the sound parameter and the data). A number of guidance systems are based on this paradigm. In [24], for example, three different frequencies (300, 600, and 900 Hz) are used to help the surgeon access the scala tympani without injuring important organs in the complex structure of the temporal bone. Wegner [4] proposed the use of amplitude modulation to guide the surgeon with a navigation system. In Scholz *et al.* [7], the authors used the pitch and the brightness to sonify a two-dimensional space and explored their effect on stroke rehabilitation. In the case of parking car systems, the distance information is provided through a decreasing time interval between impulse tones.

These two mapping paradigms, both have advantages and drawbacks. If "spatial sonification" is considered as more intuitive and natural, it is less adapted to situations where our sound localization abilities are poor, which is the case when estimating elevation or distance [25]. On the other hand, "non-spatial sonification" has proved its efficiency in many systems but generally requires a longer learning process and has a strong dependency on the auditory parameter used. Indeed, while any auditory parameter can be considered as a "display" dimension and consequently can be used for sound guidance, all the auditory parameters may not lead to the same performances.

This study focuses on the "non-spatial sonification" paradigms. It aims at exploring and quantifying the influence of several auditory parameters on a guidance task in situations where high accuracy is needed and where "spatial sonification" cues might not be available (for instance when only a single sound source is available) and aims at transmitting robust effects that could be transmitted on poor quality loudspeakers in noisy environments. As the robust comparison of several sonification strategies requires objectivity, systematicness, and reproducibility [26], this article focuses on guidance in one-dimensional space with a task that is sufficiently generic to be transposable to various types of applications. It thus presents a method that aims at identifying and comparing sound attributes for precise, rapid and direct (no overshooting) auditory guidance.

The article first proposes three types of informative sonification strategies based on two main categories to perform a guidance task. Then, it presents the design of several sonification strategies for each defined category. Finally, it presents results of a comparative perceptual evaluation of the designed strategies performed using a guidance task. The obtained results provide relevant information for prediction of the user's behavior with a chosen sonification strategy and constitute a first step toward general guidelines for mapping auditory parameters onto data dimensions.

II. SONIFICATION STRATEGIES FOR GUIDANCE TASKS

To investigate the influence of specific signal structures on guidance behavior without any specific application in mind, the concept of "relative distance" is introduced. Thus, instead of guiding the user with quantities that are specific to a given application, the sound parameter is mapped to a "relative distance" between current and target data values. This involves the definition of specific data values considered as targets, which may change over time (in the case of dynamic guidance). In the general case, the target(s) correspond to one (or several) requested system state(s) between which the user (or any process) is moving, and in which the information to sonify corresponds to the absolute value of the distance to these targets. As an example, in a driving aid application, the target may represent the optimal speed for the road section on which the user is located. The system will then give information on the velocity difference to be applied to reach the optimal speed. The data to display will then be the distance between the current and the target speed. Note that, in this case, the target may vary dynamically.

In order to address applications with different scales in an overall manner, a normalization of the distance (by the maximum data value) is proposed. Hence, the sonified data is no longer a physical dimension but a relative distance that is always dimensionless. The maximum data value is defined once and for all by the designer as a function of the required precision for a specific application and according to the set up apparatus in order to favor the process of learning once it has been defined. In all applications including a single target, the normalized distance varies between 1 (the user is at the maximum distance from the target) and 0 (the user is on the target with a margin of precision). For example, in the case of the parking car systems, the proximity sensors have a detection range of approximately 2 meters. The maximum data value is set to 2 meters and the normalized distance varies between 1 (when the obstacle is at a distance of 2 meters) and 0 (when the obstacle is at the minimal pre-defined distance). By using such a process, the same sound variations can be applied to different kinds of displacements, such as the distance (in meters) of a walking person or the rotation (in degrees) of a car's steering wheel. This process allows to compare the use of different sound strategies independently of the data. The abstraction process is thus based on the definition of one (or several) target(s) and on calculations of the normalized distance between current and target values of the data to be sonified.

Auditory display in a guidance task can have multiple goals:

- to guide as precisely as possible,
- to guide as quickly as possible,
- to guide without passing the target (e.g. presence of an obstacle or prohibited area).

These guidance goals should directly affect the choice of the sound design as some sound parameters may mainly influence rapidity, whereas others may influence precision or overshooting.

According to these 3 goals we investigated the efficiency of several sonification strategies in order to identify signal attributes that are best suited for guidance goals. In order to focus on the guidance performance of each sound attribute, basic synthetic sounds were used in this experiment. In future applications, such sounds can easily be combined with various sound textures to improve aspects related to the user satisfaction (sense of comfort, pleasure, well-being). However such considerations are beyond the scope of this study.

We introduce a categorization based on several assumptions related to the way certain sound attribute variations are expected to affect guidance efficiency (in terms of precise, rapid, and direct guidance) without a specific learning process. The categorization was constructed on the basis of the results of one of our previous studies [27], [28] which firstly introduced the sonification categories and its evaluation and highlighted the need to separate precise, rapid and direct guidance in the instructions. Indeed, the evaluation task didn't evoke any specific goal for the guidance task and the results showed different subject behaviors as a function of the weight they put on rapidity or on precision. The two sonification categories correspond to different kinds of variations of the sound attribute. The first category corresponds to the variation of the main auditory attributes (i.e. pitch, tempo, loudness, and timbre). For this category, the profile of each strategy corresponds to a simple variation of the auditory attributes. The second category contains an auditory reference corresponding to the target. This results in an auditory profile that will be characterized by a specific sound on the target. On the basis of this second main category, it is possible to define a third category characterized by the presence on the target of an auditory reference declined at several scales in order to create a zoom effect. The sound profile of this category is similar to the second but with more variations around the target.

- "*Strategies without reference*": These strategies are based on the variation of basic perceptual sound attributes such as pitch, loudness, tempo, or brightness and other timbral parameters. The sound attribute varies as a function of the normalized distance to the target between a minimum value (when the user is on the target) and a maximum value (for the maximum distance that could be reached by the user). For these strategies, it is necessary to define the polarity (i.e. whether the auditory parameter is maximum or minimum on the target), the mapping function (linear, exponential, etc.), and the range of sound parameter values. Since the extreme values are unknown to the user, he/she has no prior knowledge about what the target should sound like. Thus, this category is based on variations of the auditory parameter in a specific range. The hypothesis here is that lack of

knowledge related to the target sound will inevitably force the user to overshoot the target and to oscillate around it before he/she finds it. Furthermore, such strategies are constrained by human perceptual limits, meaning that the maximum attainable precision will probably be limited by the just noticeable difference (JND) for each sound parameter.

- "*Strategies with reference*": The idea here is to generate sounds that contain an auditory reference corresponding to the target. The target is represented by a reference sound and the distance is represented by the presence of another component that varies as a function of the normalized distance to the target and matches the reference sound on the target. With this auditory reference, the user should be able to evaluate the distance to the target at each step of the guidance process (by analyzing the instantaneous variation contained in the sound) without needing to explore the full range of the sonification strategy in advance. It should be noted that the range is less important in these strategies (the target is not necessarily defined by a minimum or a maximum value but by a specific sound) and the polarity is fixed by the relative positioning of the varying component with regards to the target. In addition, the notion of reference can be addressed as an implicit perceptual reference such as the inharmonicity [29] (the sound is harmonic on the target and becomes increasingly inharmonic as we move away from the target) or the roughness (there is no roughness on the target). With an implicit reference, the polarity is always positive as the target corresponds to the minimum value of the sound parameter (there is no inharmonicity or no roughness on the target). As the sound is different on the target, it is assumed that strategies from this category will prevent the user from overshooting the target or at least leading to fewer oscillations around the target than for the previously presented "*strategies without reference*". However, such strategies may in some cases also lead to a longer guidance time. For example, when frequency modulations are used as a strategy with modulations that decrease when approaching the target, the subjects may tend to slow down to find the exact target location.
- "*Strategies with reference and zoom effect*": It is hypothesized that it is possible to improve "strategies with reference" by adding a "zoom effect" that may increase the precision around the target and reduce the target identification time. In this case, the reference is represented by a set of several parameter values and the distance is represented by the presence of another set of auditory components, each of which varies from a maximum value to the reference values. The zoom consists of amplifying the reference effect as the target is approached by applying a multi-scale variation (i.e. a variation that differs for each component of the set) so that the target can be reached with higher precision. For this category, the sonification strategies can generally be applied to several perceptual attributes

(pitch-related, temporal, and timbral). For example, if we introduce a zoom effect in the frequency modulation strategy (belonging to the previous "strategies with reference" category), it involves the creation of additional modulations scaled at different frequency ranges, i.e. low modulation for low frequency band and high modulation for high frequency band. Hence, even if the user is close to the target, a high modulation could be heard allowing the user to still be efficiently informed on the distance. This may favor precise guidance and reduction of the identification time and, as they belong to strategies with reference, few oscillations around the target.

Given these sonification strategy categories, it is therefore possible to imagine creating a number of corresponding sonification strategies and then performing a comparative evaluation of the effect of these strategies on the user behavior during the guidance task.

III. Design of the sonification strategies within each category

In order to perceptually evaluate guidance behavior as a function of sound attributes, several sonification strategies were created using the proposed categorization. As per the description introduced in [30], sound parameters from pitch-related, temporal, loudness-related, and timbral categories were investigated. Interestingly, note that in [30] the authors reported that Loudness, Pitch, and Duration were the three most used acoustical parameters for representing distance.

Given that the implementation of a sonification strategy is not unique (for example, variation of inharmonicity can be implemented in several ways), some design choices were made: only one auditory parameter varied in the "*strategies without reference*" category to favor a single auditory stream in the generated sound. The sound parameters varied orthogonally across strategies (e.g. a pitch variation did not contain a loudness variation within a given strategy). For strategies without reference, where possible, the mapping functions took into account perceptual properties of human hearing so that the variations were perceived linearly over the whole distance. Whereas the limits of frequency perception are often quoted at 20 – 20000 Hz for young healthy listeners, the range values were here determined to cover at best the values available in everyday audio devices (i.e. from 300 to 3400 Hz).

Nine sonification strategies were created taking into account these design choices (note that this is not an exhaustive list and several other sonification strategies might be defined for each category). These strategies are described below and resulting sound examples are available online.[1]

[1]http://www.lma.cnrs-mrs.fr/~kronland/IEEE_SonificationStrategies/

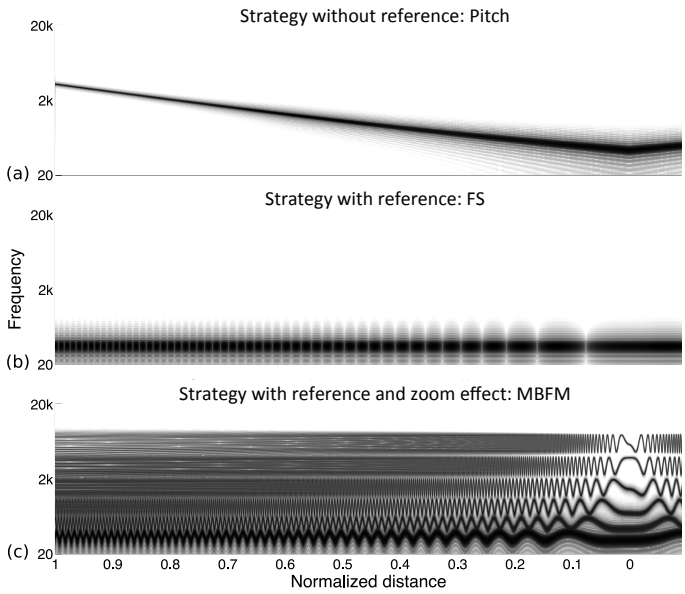


Fig. 1: Spectrograms of the sounds generated by three sonification strategies (one example for each category) simulating a varying normalized distance. (a): Pitch strategy (for Strategies without reference), (b): Fluctuation Strength (for Strategies with reference), and (c): Multi-Band Frequency Modulation (for Strategies with reference and zoom effect).

A. Strategies without reference

1) Pitch: The pitch strategy consists of mapping the normalized distance onto the frequency of a pure tone (cf. Figure 1a). This strategy is implemented using a sine wave of varying frequency $f(x)$ based on the normalized distance to the target $0 \leq x \leq 1$:

$$s(t) = A(f(x)) \cos\left(2\pi f(x) t\right)$$

As human perception of frequency varies logarithmically, we chose the following scaling function $f(x)$:

$$f(x) = f_{min}.2^{x.n_{oct}}$$

where s is the sine wave, t is the time parameter, $n_{oct} = \ln \frac{f_{max}}{f_{min}} \times \frac{1}{\ln 2}$ is the number of octaves covered by the strategy and f_{min} and f_{max} are the extreme frequency values. To normalize the variation with respect to loudness, the amplitude A of the sine wave is weighted by the isophonic curve depending on the frequency $f(x)$ from Standard ISO 226 [31]. The polarity was chosen such that the frequency was minimal on the target. The range of the scaling function was set to frequencies corresponding to traditional

telephone bandwidth (300 – 3400 Hz): $f_{min} = 300$ Hz and $f_{max} = 3394$ Hz, hence spanning 3.5 octaves.

2) Tempo: This strategy consists of mapping the normalized distance onto the repetition rate of a generated sound. Due to the temporal perception of the human ear, the polarity classically used for this strategy leads to the maximal repetition rate on the target. Thus the closer the target, the faster the sound repetition.

The sound stimulus was defined as a pulse tone of frequency $f_0 = 1000$ Hz and of duration $T = 0.1$ sec. The repetition rate was set to 2 Hz (120 bpm) for the maximum distance and changed linearly up to 20 Hz (1200 bpm) on the target.

3) Loudness: Based on loudness perception of sounds, this strategy is implemented using a sine wave of frequency $f_0 = 600$ Hz with a varying amplitude $A(x)$ based on the normalized distance to the target $0 \leq x \leq 1$:

$$s(t) = A(x).\cos(2\pi f_0 t)$$

As the relationship between sound pressure level and loudness can be approximated by a power function [32], $A(x)$ is:

$$A(x) = 10^{[(\log A_{max} - \log A_{min}).x + \log A_{min}]}$$

For consumer applications (on mobile phones, for example), the maximum available dynamic level is generally around 40 dB. For the experiment, the dynamic range was therefore limited to 40 dB and the polarity was chosen so that the loudness was minimum on the target: $a_{min} = -40$ dB with a maximal value of $a_{max} = 0$ dB such that $A_{min} = 10^{-\frac{a_{min}}{20}} = 0.01$ and $A_{max} = 1$. In order to be sufficiently general, this strategy is based on a relative variation of the level. The 40 dB range is added to the device level (L_{device}) so that the sound level at the target is 40 dB SPL weaker than at the maximum distance from the target.

Note that this strategy is not strictly orthogonal to the pitch parameter as some studies (e.g. [33]) report slight change in pitch perception with an increase of intensity. However, considering the high interindivual differences of this effect and its rather weak influence, this effect was considered as marginal.

4) Brightness: Brightness is the auditory analogy to visual brightness, and is considered to be one of the most frequently used perceptual auditory attributes related to the timbre of a sound. It corresponds to an indication of the frequency distribution contained in a sound, and is highly correlated to the spectral

centroid [34]. For this strategy, brightness variations are obtained using second order lowpass filtered white noise with a logarithmic distance-dependent cutoff frequency F_c:

$$F_c(x) = f_{min}.2^{x.n_{oct}}$$

where $n_{oct} = \ln \frac{f_{max}}{f_{min}} \times \frac{1}{\ln 2}$ is the number of octaves covered by the strategy and f_{min} and f_{max} are the extreme frequency values.

As for the pitch strategy, the range of the scaling function was set to frequencies corresponding to the traditional telephone bandwidth (300 – 3400 Hz): $f_{min} = 300$ Hz and $f_{max} = 3394$ Hz hence spanning 3.5 octaves.

With this mapping, the spectral centroid (computed with the MIR Toolbox [35]) is 700 Hz on the target and 3600 Hz at the maximum distance from the target.

B. Strategies with reference

1) Inharmonicity: According to several studies, untrained Western subjects are sensitive to the harmonicity of sounds and thus can easily detect divergence of overtone frequencies from a harmonic series. Using this ability, the present strategy is based on inharmonicity perception of sounds and uses an implicit perceptual reference: the harmonic sound. The sound is constructed using a sum of N sine waves whose fundamental frequency is $f_0 = 200$ Hz and with higher frequencies computed using the inharmonicity formula proposed by Young [36] for the piano:

$$s(t) = cos(2\pi f_0 t) + \sum_{k=2}^{N+1} \cos(2\pi f_k \sqrt{1 + b(x)k^2} t)$$

where $f_k = kf_0$, and $b(x)$ is the inharmonicity factor based on the normalized distance x such that $b(x)$ varies between 0 and 0.01 (i.e. range of values observed in piano strings [37]).

With this mapping, the inharmonicity index (computed with the MIR Toolbox [35]) is 0. on the target and 0.45 at the maximum distance from the target.

2) Fluctuation Strength: This strategy is designed by creating an explicit reference in the sound which is a pure tone of frequency $f_0 = 200$ Hz. An additional pure tone is then considered, whose frequency varies from $f_0 + 10$ Hz (for the maximum distance) to f_0 on the target. The variation of the fluctuation strength is obtained from the frequency distance between these pure tones:

$$s(t) = 0.5 * cos(2\pi f_0 t) + 0.5 * cos(2\pi(f_0 + 10x)t)$$

The result is an amplitude modulation with a frequency equal to the difference between the two tones [38]. With the chosen values, when the normalized distance x equals one, there are 10 modulations per second. When the target is reached, no more beats are heard (cf. Figure 1b). The modulation rate of the fluctuation strength is thus comprised between 0 Hz (on the target) and 10 Hz (at the maximum distance from the target). The fluctuation strength (computed with the MIR Toolbox [33]) varies between 0 and 0.34 vacil.

3) Synchronicity: This strategy is an extension of the tempo strategy and is based on the repetition of two identical sounds. The first sound corresponds to the reference, the second is time-shifted with a varying delay time Δt based on the distance to the target. The sound is constructed using a pulse of a harmonic sound of fundamental frequency $f_0 = 400$ Hz with nine harmonics, an attack time of 5 ms and a release time of 495 ms. The repetition rate is set to 2 Hz (120 bpm). When the distance is maximum, the second pulse is shifted by 1/4 of the pulsation frequency (e.g. 125 ms). When the target is reached, the two pulses are synchronized.

C. Strategies with reference and zoom effect

1) Multi-Band Frequency Modulation (MBFM): This strategy is based on frequency modulation of a harmonic sound of fundamental frequency $f_0 = 200$ Hz. Here, each harmonic is frequency modulated in a different way: the modulation frequency of the k^{th} component is $f_m(x) = 10k.x$ and depends on the normalized distance x such that the higher the frequency of the component, the higher the frequency of the modulation signal. When the user approaches the target, the modulation frequency decreases (there is no modulation when the target is reached). The further the target, the higher the modulation frequency and the more complex the sound:

$$s(t) = \sum_{k=1}^{N} sin(2\pi f_k t + Ik.sin(2\pi f_m(x)t))$$

where f_k is the frequency of the k^{th} harmonic, $I = 50$ is the modulation index, and $f_m(x)$ is the modulation frequency.

Use of a harmonic sound allows construction of an "auditory zoom". The concept is simple: frequency modulation affects all harmonics but with different temporalities. For a fixed distance, the higher the frequency, the faster the modulation. Near the target, the modulation frequency of the first harmonic is too small to rapidly grasp the target, but the modulations from the second harmonic, which is twice as

fast, and then from the third harmonic (three times faster) enables faster and more precise location of the target (cf. Figure 1c).

2) Multi Scale Beating (MSB): This strategy is based on the same concept of auditory zoom as the MBFM strategy. It uses a sound of N harmonics of fundamental frequency f_0 duplicated M times. The mth duplicated spectrum is transposed by a factor $m(\alpha(x)-1)$ that depends on the distance x to the target. The strategy is constructed on a multi-scale variation and then defined by:

$$s(t) = \sum_{k=1}^{N} \sum_{m=0}^{M} A_k \cos(2\pi f_k (1 + m(\alpha(x) - 1))t)$$

where $0.94 \leq \alpha(x) \leq 1.06$ and $N = 15$, $M = 11$ and $f_k = kf_0$ with $f_0 = 200$ Hz. Similarly to the MBFM strategy, the "auditory zoom" is due to the use of a harmonic sound and to a modulation frequency that depends on harmonic order. The MBFM and MSB strategies differ by the fact that the modulation is based on frequency for the MBFM strategy and is temporal for MSB strategy.

IV. METHOD

An experiment was designed to explore the ability of previously defined sonification strategies (see section III) to dynamically guide the user toward a hidden target. For that purpose, subjects were asked to perform a hand guidance task on a graphic tablet and the protocol was restricted to a one-dimensional, one-polarity task. This experiment first aimed at quantifying the efficiency of the designed strategies in terms of precision, rapidity, and displacement around the target. Then, it tried to quantitatively assess the behavioral differences in the motor control tasks induced by the different strategy categories.

A. Subjects

Twenty-four voluntary subjects participated in the experiment (6 women and 18 men; mean age: 25.0 $\pm$ 3.4 years (min. 20; max. 32 years)). All were naive regarding the purpose of the experiment and none of the subjects reported any hearing losses. Twenty-two subjects self-reported as right handed, and two as left handed.

B. Stimuli and Apparatus

The subjects were placed in a quiet room wearing stereo closed-ear headphones (model Sennheiser HD280). They sat in front of a graphic tablet (model Wacom Intuos 5 - active surface: 325.1 mm x 203.2 mm - acquisition data rate: 133 Hz - precision with standard nib: 5x10^{-3} mm) and a computer screen. They were placed so that the pen tablet was beside their dominant hand and the computer

keyboard next to their other hand. The experiment ran on an interactive interface implemented using the Max programming environment[2] and presentation of the instructions was automated (pre-recorded vocal messages).

The sound stimuli were synthesized in real-time using the nine strategies defined in section III. In each strategy, a sound parameter changes as function of the normalized distance between pen and hidden target. The overall sound level was fixed at a comfortable level at the beginning of the experiment and subjects were not able to modify it during the experiment.

To avoid a potential learning effect regarding the target tone in the pitch strategy, the target frequency for the *pitch strategy* was randomly selected between 200, 250, 300, and 350 Hz while maintaining a range of 3.5 octaves. For the other strategies, randomization of the target value was not necessary as the memorization process for loudness, brightness, harmonicity or tempo (near 1200 bpm) was judged not to influence the subject.

C. Procedure

The participants were told to find a hidden target randomly placed on a virtual horizontal line on a pen tablet (user movements were not constrained, thus the target corresponded to a hidden vertical line). The starting position was the same for all participants and trials and was defined on the left border of the tablet. From this starting position, four physical target distances were considered: 20, 22.5, 25, and 27.5 cm. Hence, the maximum attainable distance by the subject in this task was 27.5 cm. Thus, the four target distances were normalized with respect to this distance leading to normalized distances of 0.73, 0.82, 0.91, and 1.

For each trial, subjects first placed the pen on the starting position. Then, they launched the trial by pressing the space bar of the computer keyboard and explored the virtual horizontal line while listening to the actual sonification strategy. Once they believed they had found the hidden target with the pen, they validated the final pen position by pressing the space bar of the keyboard. Validation automatically triggered the following trial and subjects had to return to the starting position to begin the new trial. Subjects were instructed not to remove the pen from the tablet during the trial. No accuracy feedback was provided.

The whole experiment was divided in three separate sessions in which the following instructions were specified to find the hidden target: “be as precise as possible” for the first session, “be as quick as

[2] http://cycling74.com/downloads/

possible" for the second session and "don't overshoot the target" for the third session. The nine strategies were evaluated within each session in order to explore the sound induced behavior as a function of the instruction. Hence, each session contained nine blocks (for the nine strategies) of four trials (each block contained each of the four target distances). Sessions, blocks and trials were presented in random order to avoid any potential learning effect. The random order of the sessions was calculated using a Latin square design in an attempt to counterbalance any potential learning effects.

At the beginning of each session, subjects were informed of the instruction to follow (i.e. "be as precise as possible", "be as quick as possible", or "don't overshoot the target") by a pre-recorded vocal message. Then, at the beginning of each block, a vocal description of the sonification strategy (the sound parameter that represented the distance) was given[3]. Subjects could then explore this strategy with a training trial in which the user position and the target were visually located on the horizontal line (placed at a different position than that of the four targets in the "real" trials). The points corresponding to the visual target and the user position had a diameter of 0.25 mm. With this training, subjects were able to understand the sonification strategy and to familiarize themselves with the apparatus. After the training session, the four trials associated with this strategy were presented. No feedback was given to the subject regarding the four target positions.

The trials of the session corresponding to "be as quick as possible" lasted a maximum of 5 seconds (from the moment the subject pressed the space bar). After these five seconds, the sound automatically stopped and the next trial was presented. This cutoff was chosen in order to force the subject to perform the task as quickly as possible. Several pretests revealed that the notion of rapidity varied strongly between subjects. This time constraint of 5 seconds enables comparison of all subjects' results on the same basis. In order to familiarize the subjects with such a quick task, a training block was proposed, instead of a single training trial with a sonification strategy, that differed from the nine evaluated strategies[4]. This training block was composed of a familiarization trial (with visual display) and as many test trials as the subjects needed to get used to the rapidity of the task (without visual display). In general, subjects needed a mean of $7(\pm 2)$ trials to be at ease with the task. This training block was not necessary for the two other sessions as they did not contain any time restriction. After this training block, the experimental procedure of this session was exactly the same as for the two other sessions (nine blocks each consisting

[3]All the descriptions are available on the website: http://www.lma.cnrs-mrs.fr/~kronland/IEEE_SonificationStrategies/

[4]This strategy belongs to the strategies with reference and the zoom effect category. It is constructed in the same way as the MBFM strategy but with amplitude modulations instead of frequency modulations.

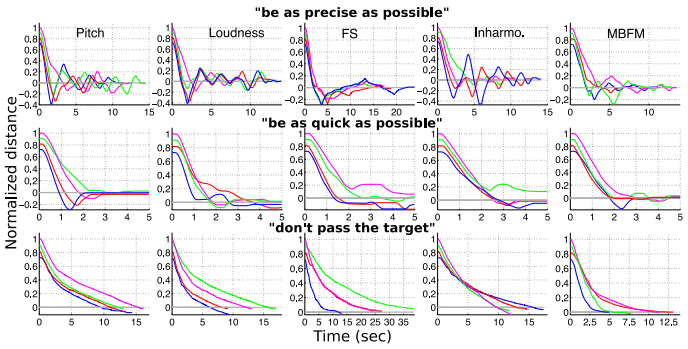


Fig. 2: Examples of results for one subject for the three sessions, for five of the nine strategies (one graph per strategy). Instruction "be as precise as possible" on top, "be as quick as possible" in the middle, and "don't overshoot the target" on the bottom line. Each graph represents the pen/target normalized distance as function of time (in seconds, i.e. note that scales are not the same) for the four distances (0.73 in blue, 0.82 in red, 0.91 in green, and 1. in magenta).

of a strategy familiarization trial followed by four trials).

Except for the strategy familiarization trials, all possible visual cues that could be exploited by the subject were removed from the screen and the tablet.

D. Data analysis

For each subject and each trial, the final error (i.e. the final absolute normalized pen/target distance on the horizontal axis, as a percentage), and the identification time (calculated between the launch and validation event triggered by the subject) were computed. Note that the vertical deviation of the pen was not taken into account for the analysis. Hence the pen/target distance was computed based on the projection of the pen and target positions along the horizontal axis.

Figure 2 gives representative tendencies for the subjects' behaviors. It shows results obtained by one subject for each session and for five of the nine strategies. The distance to the hidden target is represented as function of time for the five strategies and all trials. Based on observations of subjects' results, specific analyses were conducted depending on the session. For the session that focused on precision, a large number of oscillations around the target were observed. These oscillations were represented by the number of crossings of the horizontal time-axis (number of zero crossings). Subjects also had a tendency to interrupt their movement to listen to the sound feedback. These interruptions were taken into account when exceeding 250 ms and the downtime representing the sum of all interruption times in one trial, was

calculated. To examine the subjects' behaviors under rapidity constraints, two phases were distinguished in the identification time: an approach time, representing the time between the beginning and the first direction change of the pen along the x-axis after having passed the target and an adjustment time, corresponding the remaining time that lasted until the end of the trial. Finally, for the session in which subjects were instructed to not overshoot the target, the passing rate (number of trials in which the subject passed the target) was calculated as a function of the overrun.

For each session, these descriptors were averaged across the four trials (distances) for each subject and analyzed in a one-way repeated measures analysis of variance (ANOVA) with "strategy" as the within-subject factor (9 levels). For all statistical analyses, effects were considered significant if the p-value was less than or equal .05. All p-values were adjusted (using a Bonferroni correction) for multiple testing (post-hoc tests).

V. RESULTS

A. Session "be as precise as possible"

Results for each strategy are presented in Figure 3. The mean value of the final error over all the strategies is $0.97\% \pm 0.43\%$ (approximately 2.7 mm). Analysis of the performances highlights a strong effect of strategy [$F(8, 184) = 43.8$, $p < 10^{-5}$]. The lowest errors are obtained with the "*strategies with reference and zoom effect*" (*MBFM* and *MSB*) with a mean error around 0.1% and 0.2% (approximately 0.27 mm and 0.55 mm). Pairwise post-hoc comparisons reveal significant differences between these two strategies and the others ($p < 0.001$). For "*strategies without reference*", performances depend on the sound parameter used. Errors obtained with *pitch* and *tempo* strategies are similar and around 0.5% whereas *loudness* and *brightness* strategies lead to the highest errors and larger inter-subject differences. Post-hoc analysis highlights significant differences between *pitch/tempo* and both *loudness/brightness/inharmonicity* and *synchronicity* ($p < 0.001$). For "*strategies with reference*", the lowest errors are obtained with the *FS* strategy, whereas the highest errors are found for *synchronicity* (with intermediary results for the *inharmonicity* strategy). Post-hoc analysis reveal significant differences between each of these three strategies ($p < 0.001$).

With regard to the total time spent on each trial in this session (figure 3b), on average, subjects took 21.6 ± 3.8 seconds to find the target. Analysis highlights a significant strategy effect [$F(8, 184) = 4.42$, $p < 0.001$]. Post-hoc tests show significantly faster responses for the *pitch* strategy compared with the *brightness/FS/synchronicity* and *MSB* strategies ($p < 0.05$).

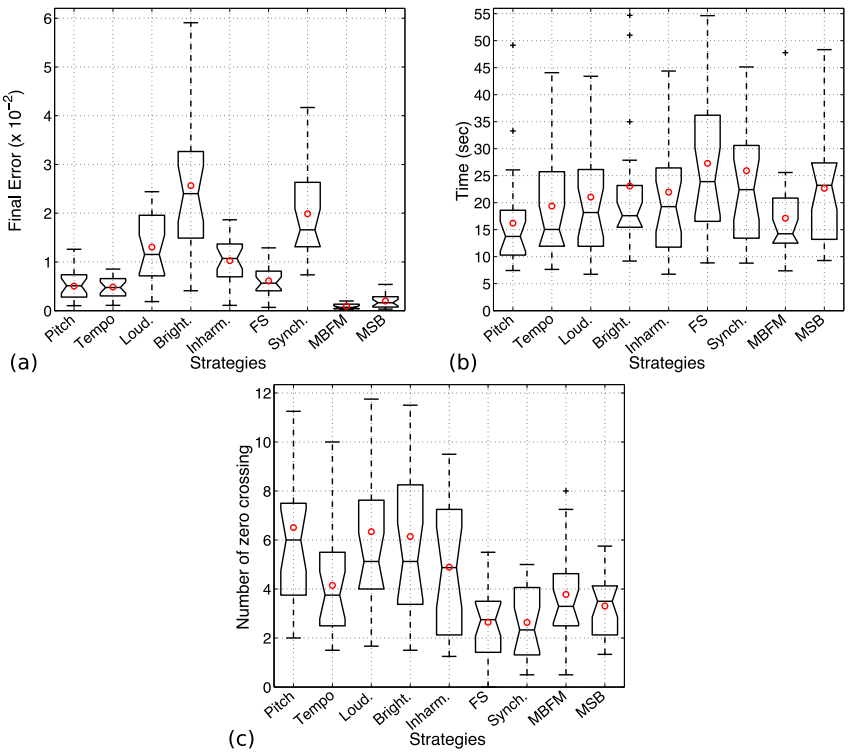


Fig. 3: Session "be as precise as possible": boxplots of final error (a), identification time (b), and number of zero crossings (c) for all subjects as a function of the sonification strategy. Mean values are represented by a red circle.

The mean number of interruptions across all strategies is 11.9 ± 3.4. Statistical analysis revealed a significant effect of strategy for downtime [$F(8, 184) = 7.85$, $p < 0.001$]. As expected, the *FS* strategy induced longer downtime (12.3 ± 9.5 sec) than the other strategies, in particular "*strategies without reference*" as well as the *inharmonicity* and *MBFM* strategies.

The number of zero crossings (figure 3c) provides information on the behavior induced by the strategy type. Indeed, the main result of the statistical analysis on this descriptor [$F(8, 184) = 9.53$, $p < 10^{-5}$] revealed that the "*strategies without reference*" and the *inharmonicity* strategy induced more oscillations around the target than the other strategies. In particular, significant differences were found between the *FS/synchronicity* strategies and the *pitch/loudness/brightness/inharmonicity* strategies ($p < 0.05$).

B. Session "be as quick as possible"

Results are illustrated in Figure 4. During this session, the overall performance in terms of final error is $3.68\% \pm 0.68\%$. Analysis of the performances highlights a significant strategy effect [$F(8, 184) =$

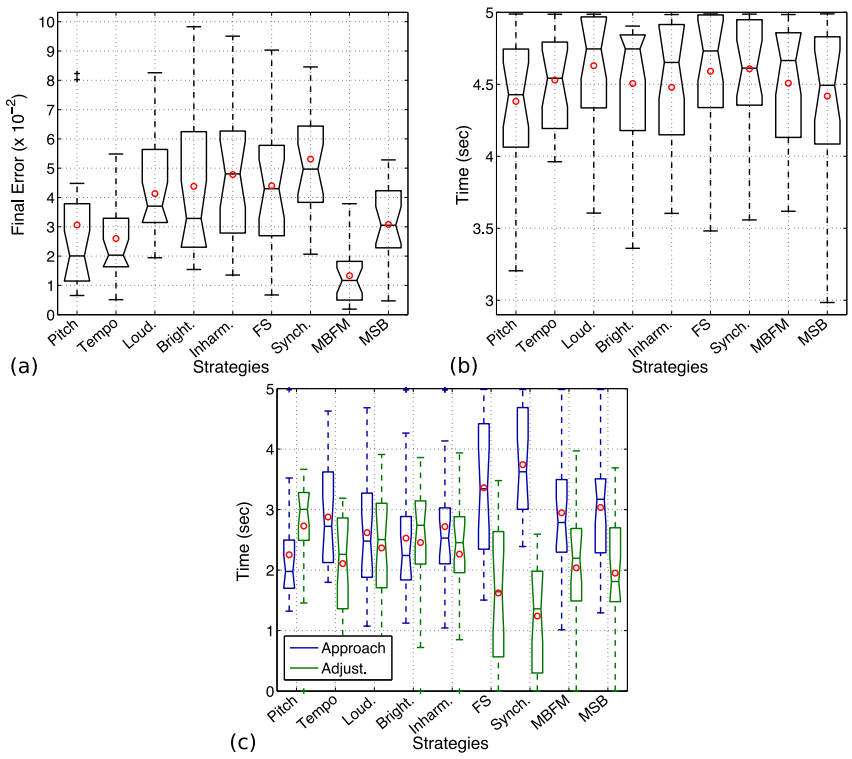


Fig. 4: Session "be as quick as possible": boxplots of the final error (a), identification time (b), and approach (in blue) and adjustment (in green) times (c) for all subjects as a function of the sonification strategy. Mean values are represented with a red circle.

9.58, $p < 10^{-5}$]. The lowest errors are obtained with the *MBFM* strategy and the highest errors with the *synchronicity* strategy (see figure 4a). Post-hoc analysis revealed significant differences between the *MBFM* strategy and all the other strategies except the *pitch* and *tempo* strategies ($p < 0.001$) and significant differences between the *tempo* strategy and the *loudness/FS/synchronicity* strategies ($p < 0.05$).

With regard to the total time spent on each trial, on average, subjects took 4.5 ± 0.1 sec to find the target. The data distribution shows that, for all strategies (except for the *FS* strategy), 75% of trials were validated before the end of the five seconds (the boxplots' third quartiles are between 4.7 and 5 seconds). Regarding the mean identification time, no statistical differences were found between strategies [$F(8, 184) = 1.18$, $p = 3.15$]. Analyses of approach and adjustment times revealed different tendencies: the approach time was shorter than the adjustment time for the *pitch* strategy, whereas it was similar to the adjustment time for the *loudness*, *brightness* and *inharmonicity* strategies, a bit larger for the *tempo*, *MBFM* and *MSB* strategies, and much larger for the *FS* and *synchronicity* strategies. Statistical analyses

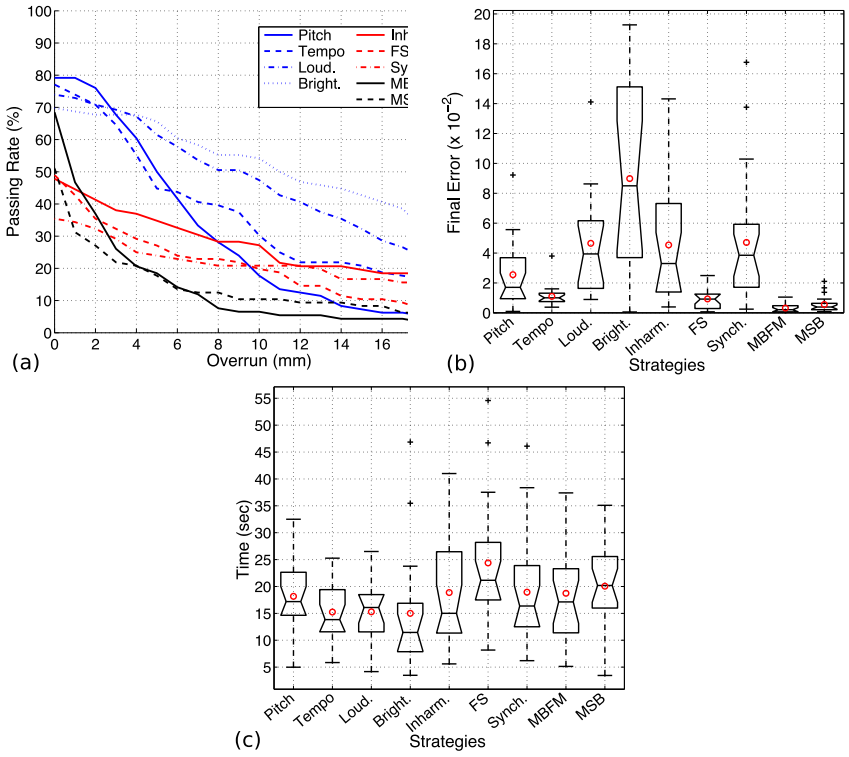


Fig. 5: Session "don't pass the target": passing rate as a function of overrun (a), boxplots of the final error (b), and of identification time (c) for all subjects as function of the sonification strategy. Mean values are represented with a red circle.

confirmed these tendencies and highlighted a strategy effect on the approach time [$F(8, 184) = 11.04, p < 10^{-5}$]; significant differences were found between the *pitch* strategy and the *FS/synchronicity/MBFM/MSB* strategies ($p < 0.05$), and between the *synchronicity* strategy and both the "*strategies without reference*" and *inharmonicity* strategies ($p < 0.001$).

C. Session "don't pass the target"

Figure 5a presents the passing rate of the target as a function of the overrun (the maximum distance from the target when the subjects overshoot the target). Changes in the passing rate as function of the overrun highlighted differences between strategies. For a 0 mm overrun value, the target was overshot in 75% of the trials when using the "*strategies without reference*" (blue curves), in 44% of the trials when using the "*strategies with reference*" (red curves) and in 60% of the trials when using the "*strategies with reference and zoom effect*" (black curves). This means that the subjects were able to detect an overrun more easily with strategies with reference than with "*strategies without reference*". Since very few trials

had a 0 mm overrun value, we decided to use a threshold at 5 mm overrun to keep more than half of the trials. The final error (figure 5b) and the identification time (figure 5c) were calculated based on these trials. Hence, the overall performance in terms of final error was $3.15\% \pm 2.14\%$. Note that this error is larger than the 1.8% (corresponding to the normalized overrun 0.5/27.5) obtained in the trials in which subjects stopped more than 5 mm before the target.

Analyses highlighted a strong effect of strategy on the final error [$F(8, 184) = 13.1$, $p < 10^{-5}$] with lower performances (higher errors) for the *brightness* strategy compared to the other strategies ($p < 0.05$) and better performances for the "*strategies with reference and zoom effect*" compared to *pitch/loudness/brightness/inharmonicity/synchronicity* strategies.

The mean identification time was 16.5 ± 1.2 sec. No significant effect of strategy on identification time was found for this session [$F(8, 184) = 1.1$, $p = 0.62$].

VI. Discussion and general guidelines for sonification design

As per the proposed categorization of sonification strategies based on specific variations of sound attributes for the purpose of guidance task, three categories of strategies (strategies without reference, "*strategies with reference*" and "*strategies with reference and zoom effect*") were proposed and investigated. As expected, the results of this experiment confirmed that user behaviors and performances (in terms of precision, time spent to reach the target and number of oscillations around the target) generally depended on the instruction and on the type of sonification strategy. We discuss these two aspects in the following paragraphs.

Firstly, the results revealed different performances in the guidance task with respect to the instruction (session). Hence, as expected, the final error (i.e. distance between perceived and actual target) was lower when the subjects were asked to be precise than when they were asked to be fast or not to pass the target. For these latter two instructions, we assumed that subjects were stressed either by temporal (for "be as fast as possible") or spatial (for "don't pass the target") constraints, leading to higher errors. On the other hand, the relative performances between each strategy obtained with the instruction "don't pass the target" were similar to the relative performances obtained with the instruction "be as precise as possible", whereas they were different with the instruction "be as quick as possible" (lower performance differences were observed between "*strategies with reference*" and loudness and brightness strategies). For the instructions "be as precise as possible" and "don't pass the target" the identification times were similar (around 20 sec).

Secondly, the results revealed different performances with respect to sonification strategies. The "*strategies without reference*" induced significantly more oscillations around the target than the other strategies (cf. session "be as precise as possible") and obliged the subjects to pass the target more often to detect it (cf. session "don't pass the target"). The performances obtained within this category were highly dependent on the type of sound attribute. The *pitch* and *tempo* strategies enabled higher precision than *loudness* and *brightness* strategies, and more rapid target detection than the other strategies from this and the other categories. In particular, in the session "be as quick as possible", *pitch* was the only attribute that induced a significantly shorter approach time than adjustment time compared to all other strategies (from this and the other categories). This strategy is therefore a good candidate when rapidity is needed for a given task. As hypothesized, results obtained with the "*strategies without reference*" highlighted the JND dependency of the final errors (especially in the session "be as precise as possible"). For *pitch*, a mean error of 0.5% (i.e. frequency variation of 3.6 Hz) was found. Considering the protocol of this experiment this resolution was of the same order of magnitude as the perceptible threshold for frequency variations below 500 Hz (equal to 2 Hz according to [38] or to 1 Hz according to [39]). For *tempo*, an error of 0.5% was found, i.e. a variation of 2.25 ms for an Inter-Onset Interval (IOI) around 50 ms. This value is in line with the literature (for a 1000 Hz sine wave) that gives a just noticeable variation of 1.54 ms at 67 ms [40] and of 12.5 ms at 50 ms [41]. For *loudness*, an error of 1.3% was found, i.e. variation of 0.52 dB, which is consistent with [38], i.e. a just noticeable variation of 1 dB for a 1 kHz sine wave. Finally, for *brightness*, a mean error of 2.57% (i.e. variation of 20 Hz) was found and was in line with [38] (variation between 30 and 50 Hz for in cut off frequency for low-pass noise around 1 kHz). In summary, the final errors related to the sound attributes found in this study were in general not very far from the JND found in psychoacoustic studies. Nevertheless, considering the differences between the experimental conditions of the present experiment and psychoacoustic experiments found in the literature, these JND comparisons should only be considered for qualitative purposes. Note that the use of psychoacoustic models in sonification has already been proposed by Ferguson *et al.* [42] who presented a theoretical implementation of the psychoacoustic definition of pitch, loudness, roughness and brightness. In this context, it might be of interest to consider the JND to predict the efficiency of a given sonification strategy with a defined range of variation. For example, the *loudness strategy* was designed to have a linear mapping between the distance and the loudness in decibel. The range of 40 dB, led to a maximum precision of 1.25% which corresponds to 0.5 dB. Thus, this range must be multiplied by 2.6 to obtain the same precision as with *pitch* or *tempo* (e.g. 0.5%), which is not feasible, since it implies a range of 100 dB.

The "*strategies with reference*" category contained three strategies (*inharmonicity*, *synchronicity* and *FS*). As hypothesized, the inclusion of a reference reduced the number of oscillations and improved the ability of the subjects to not overshoot the target, as shown by a lower passing rate for these three strategies compared to "*strategies without reference*". In terms of precision, the results obtained with the *inharmonicity* and *synchronicity* strategies were similar to the results obtained with the *loudness* and *brightness* strategies for all sessions, whereas results obtained with the *FS* strategy differed across sessions. The final errors for this strategy for the sessions "be as precise as possible" and "do not pass the target" were similar to those of the *tempo* and *pitch* strategies, whereas it gave larger errors for the session "be as quick as possible". Interestingly, differences were found between strategies based on an explicit reference (*FS* and *synchronicity*) and an implicit reference (*inharmonicity*). Indeed, the number of zero crossings was significantly lower (in the session "be as precise as possible") and the approach time was longer (in the session "be as quick as possible") for strategies with an explicit reference than for strategies with an implicit reference. With these instructions, the *inharmonicity* strategy induced similar results to those of the "*strategies without reference*".

Finally, the "*strategies with reference and zoom effect*" category, which contained two strategies (*MBFM* and *MSB*) provided good performances across sessions. The presence of a reference reduced the number of oscillations as also hypothesized for the "*strategies with reference*" category (similar results as for the *FS* and *synchronicity* strategies). Due to the zoom effect, these strategies led to the highest precision among all the proposed strategies except for the session "be as quick as possible" in which the *MSB* strategy led to slightly less precision than the *tempo* and *pitch* strategies. While the identification time did not differ from the "*strategies with reference*" category, the approach time was reduced compared to the *FS* and *synchronicity* strategies, suggesting that the zoom effect tended to improve guidance. More generally, it is possible to conclude that with these strategies, users were almost able to find the target without passing it while maintaining good precision (the target was overshot with 5 mm overrun in only 20% of the trials).

With regard to the categorization, the results enabled verification of a certain number of hypotheses on the efficiency of certain sound attributes depending on guidance type (precise, rapid or without passing the target). First, the "*strategies without reference*" generally induced more oscillations around the target than the other strategies and the errors seemed to depend on the perceptual limits of the ear. Second, the "*strategies with reference*" reduced target overrun, but induced longer guidance time than "*strategies without reference*". The precision obtained with these strategies is highly variable and is not always better than the precision obtained with "*strategies without reference*". Third, the "*strategies with reference and*

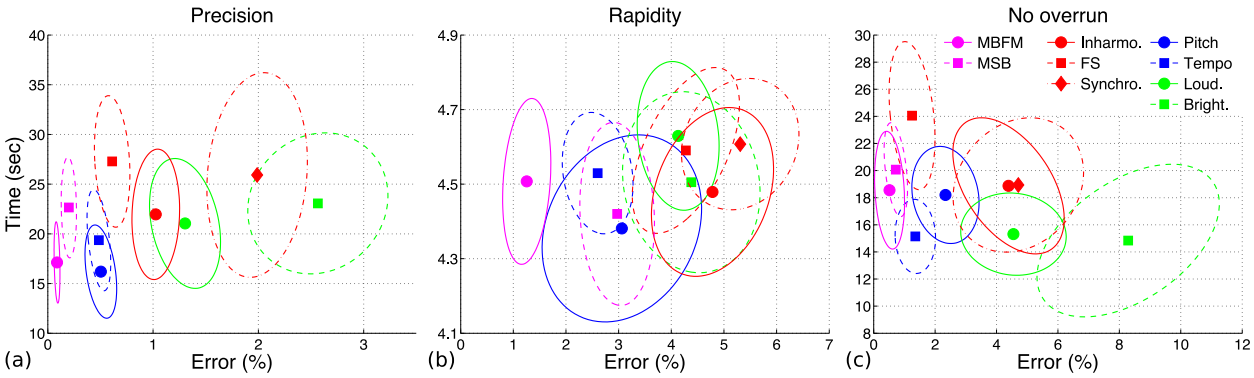


Fig. 6: Comparisons between sessions (precision (a), rapidity (b), and no overrun (c)) and sonification strategies in terms of error and time performances. Dots represent the mean values of the results obtained for error and identification time for each strategy. Ellipses represent the error distribution (scaled by a factor 4 for readability).

zoom effect" improved the precision (for each instruction), reduced the number of oscillations around the target, and almost made it possible to find the target without passing it.

The present results also allow to deeper understand the results of previous studies on guidance tasks. For example, in [7], the authors explored the effectiveness of sonification in stroke rehabilitation with the sonification of participants' computer mouse movement. With a two parameter-to-axis-mapping using pitch and brightness, their study highlighted more precise results with pitch than brightness. This results is consistent with the current study's results and could have been predicted by the present study (by analysing the position of each strategy in the error/spent time space of Figure 6).

By evaluating these sonification strategies within a single experiment and for different instructions, it is now possible to highlight some guidelines for sonification design that aim to predict user behaviors for a given guidance task. In particular, the performances obtained with the different sonification strategies can be exploited and regarded in terms of compromises between errors and acceptable time spent on task performance. Figure 6 represents performances obtained with the nine strategies within an error/spent time space as function of the instruction (a figure presenting the performances within an error/spent time/oscillation space for the instruction "be as precise as possible" is also available on the website[5]). Figure 6 suggests that the *MBFM/MSB/pitch/tempo* strategies would produce good performances independently of the type of task, whereas performances obtained with "*strategies with reference*" might

[5]http://www.lma.cnrs-mrs.fr/~kronland/IEEE_SonificationStrategies

be highly task dependent. With the help of such a space, a sound designer may be able to choose a specific sonification strategy with regard to the task, based on the predictions of guidance performance. For example, *MBFM* is, on average, the best strategy for precise and rapid guidance, whereas, if there is no time constraint, the *FS* strategy is a good candidate for a task that requires high precision and no overrun. Finally, these guidelines are of interest for proposing adjustments to a given sonification strategy to improve performances. It is, for instance, possible to adjust the mapping function between the normalized distance and the sound parameter to increase the obtained precision.

VII. Conclusions and perspectives

The aim of the current study was to compare the efficiency of several sonification strategies for guidance tasks. Three categories of sonification strategies based on two types of variation of the sound attributes were introduced. Using this categorization, nine sonification strategies were designed by sound synthesis. A perceptual experiment based on a guidance task on a pen tablet toward a hidden target was then conducted to evaluate these strategies with different instructions ("be as precise as possible", "be as quick as possible", and "don't pass the target"). Systematic investigations were performed on the link between sound attribute variations and auditory guidance efficiency, which to our knowledge has never been done before.

The results highlighted important differences between sonification strategies in terms of precision, guidance time, and oscillations around the target. Based on the results of this experiment, an "efficiency" space is further proposed as a function of the desired guidance type, aiming at informing the sound designer regarding the choice of an optimal sonification strategy in terms of rapidity, precision or target overrun. Such representations might also be useful for suggesting how efficiency can be improved for certain strategies (by changing the parameter range where possible or by changing the linear scale to a more appropriate one).

The obtained results allow predicting user performances independently of the application (but depending on the task). Therefore, the proposed guidelines for sonification design could be addressed to interaction designers in various domains involving guidance, spanning from pedestrian navigation to positional guidance in surgery.

Several perspectives can be addressed following this study. First, deeper investigations should be performed on the type of reference to be designed (explicit or implicit). In fact, categorization of the strategy with an implicit reference into "*strategies with reference*" or "*strategies without reference*" depends on the type of task (precise, rapid, or no overshooting). It might therefore be interesting to

differentiate explicit and implicit reference categories.

As this study focused on the comparison of the efficiency of several sonification strategies, the aesthetics aspects of the strategies were considered out of scope and the strategies were constructed with laboratory sounds (e.g. pure sine waves) that could be considered as irritating or annoying in everyday use. Improving user satisfaction of the proposed sonification strategies can be envisaged by applying the attribute variation of the sound strategies to more complex sounds (e.g. instrumental sounds or richer synthesized sound). As certain works have highlighted, it is indeed possible to apply variations of a perceptual sound parameter to complex sound textures without affecting sonification performances [43]. Hence, a simple design process for the sonification of guidance tasks could be: define one or several targets according to the application; select a sonification strategy that corresponds to the required guidance (precision, rapidity or passing) according to the proposed efficiency space; and apply the variation of the chosen sound parameter to complex, preferably stationary sounds.

Finally, the current study provides relevant information for predicting the user performance with a chosen sonification strategy and constitutes a first step toward general guidelines for mapping data onto auditory display dimensions and toward the identification of efficient perceptual sound structures (which are known as invariants [44], [45]) for guidance tasks. The proposed categorization was constructed for one dimensional guidance tasks and the present results are of interest mostly for this type of tasks (e.g. detect obstacles distances, represent the distance between the tip of the needle and an organ in surgery, etc.). To address additional applications, this categorization might be extended to two and three dimensional guidance tasks. This implies the analysis of the perceptual effect of combined sonification strategies as a function of the category (i.e. without reference, with reference, and with reference and zoom effect) and as a function of the type of sound attribute (based on frequency variations, temporal variations or both). This extension from one to two or three dimensions also raises a number of questions. For example, in [7], the authors use two parameters based on frequency variation (e.g. pitch and brightness) to guide the user on a 2D space. Would it be more efficient to use a parameter based on frequency variations (e.g. the pitch) and a parameter based on temporal variations (e.g. the tempo)? It would also be of interest to compare the use of one sound strategy applied to two sound streams (that represent the two dimensions) to the use of two different sound strategies applied on the same sound stream.

ACKNOWLEDGMENT

This work was funded by the French National Research Agency (ANR) under the SoniMove: Inform, Guide and Learn Actions by Sounds project (ANR-14-CE24-0018-01).

REFERENCES

[1] J. Loomis, R. Golledge, and R. Klatzky, "Navigation system for the blind: auditory display modes and guidance," *Presence: Teleoper. Virtual Environ.*, vol. 7, pp. 193–203, 1998.

[2] J. Wilson, B. Walker, J. Lindsay, C. Cambias, and F. Dellaert, "Swan: System for wearable audio navigation," in *Proceedings of the 2007 11th IEEE International Symposium on Wearable Computers*. Washington, DC, USA: IEEE Computer Society, 2007, pp. 1–8. [Online]. Available: http://dl.acm.org/citation.cfm?id=1524303.1524851

[3] G. Parseihian, C. Jouffrais, and B. F. Katz, "Reaching nearby sources: comparison between real and virtual sound and visual targets," *Front. Neurosci.*, vol. 8, no. 269, 2014.

[4] K. Wegner, "Surgical navigation system and method using audio feedback," *Proceedings of ICAD'98*, vol. 6, 1998.

[5] C. Hansen, D. Black, C. Lange, F. Rieber, W. Lamadé, M. Donati, K. Oldhafer, and H. Hahn, "Auditory support for resection guidance in navigated liver surgery." *The international journal of medical robotics+ computer assisted surgery: MRCAS*, vol. 9, no. 1, p. 36, 2013.

[6] M. Dozza, L. Chiari, and F. B. Horak, "A portable audio-biofeedback system to improve postural control," in *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, vol. 2. IEEE, 2004, pp. 4799–4802.

[7] D. Scholz, L. Wu, J. Pirzer, J. Schneider, J. Rollnik, M. Grossback, and E. Altenmuller, "Sonification as a possible stroke rehabilitation strategy," *Front. Neurosci.*, vol. 8, no. 332, 2014.

[8] A. Nyman, "Games with sounds: the blind navigation and target acquisition," in *Proceedings of seminar: Alternative Access: Feelings & Games'05. University of Tampere, Finland*, 2005.

[9] P. Eslambolchilar, A. Crossan, and R. Murray-Smith, "Model-based target sonification on mobile devices," in *Proceedings of the Int. Workshop on Interactive Sonification Organisation, Bielefeld, Germany*, 2004.

[10] A. Godbout and J. E. Boyd, "Corrective sonic feedback for speed skating: A case study," in *Proceedings of the 16th International Conference on Auditory Display*, 2010.

[11] S. Barrass, "Auditory information design," Ph.D. dissertation, 1998.

[12] K. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, and J. Neuhoff, "Sonification report: Status of the field and research agenda," Faculty Publications, Department of Psychology, 1999.

[13] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press, 1990.

[14] S. McAdams and E. Bigand, *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford: Oxford University Press, 1993.

[15] B. C. J. Moore, *An Introduction to the Psychology of Hearing (4th ed.)*. San Diego, Calif.: Academic Press, 1997.

[16] R. J. Zatorre, J. L. Chen, and V. B. Penhune, "When the brain plays music: auditory–motor interactions in music perception and production," *Nature Reviews Neuroscience*, vol. 8, no. 7, pp. 547–558, 2007. [Online]. Available: http://dx.doi.org/10.1038/nrn2152

[17] S. Gibet, *Musical Gestures: Sound, Movement, and Meaning*. Routledge, 2009, ch. Sensorimotor control of sound-producing gestures, pp. 212–237.

[18] T. Hermann and A. Hunt, "Guest editors' introduction: An introduction to interactive sonification," *IEEE MultiMedia*, vol. 12, no. 2, 2005.

[19] A. Hunt and T. Hermann, "Interactive sonification," in *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds. Logos Publishing House, 2011.

[20] A. Effenberg, "Movement sonification: Effects on perception and action," *IEEE MultiMedia*, vol. 12, no. 2, 2005.

[21] T. Lokki and M. Grohn, "Navigation with auditory cues in a virtual environment," *IEEE MultiMedia*, vol. 12, no. 2, 2005.

[22] D. Begault, *3-D Sound for Virtual Reality and Multimedia*. Cambridge: Academic Press, 1994.

[23] G. Parseihian and B. Katz, "Morphocons: A new sonification concept based on morphological earcons," *Journal of the Audio Engineering Society*, vol. 60, no. 6, pp. 409–418, 2012.

[24] B. Cho, N. Matsumoto, S. Komune, M. Hashizume, and N. Matsumoto, "Surgical navigation system for guiding exact cochleostomy using auditory feedback: A clinical feasibility study," *BioMed Research International*, 2014.

[25] J. Blauert, *Spatial Hearing, The Psychophysics of Human Sound Localization*. Cambridge: MIT Press, 1997.

[26] T. Hermann, "Taxonomy and definitions for sonification and auditory display," in *Proceedings of the 14th International Conference on Auditory Display, Paris, France*, 2008.

[27] G. Parseihian, C. Gondre, M. Aramaki, R. Kronland-Martinet, and S. Ystad, "Exploring the usability of sound strategies for guiding task: toward a generalization of sonification design," in *Proceedings of the 10th International Symposium on Computer Music Multidisciplinary Research, Marseille, France, 15-18 Oct. 2013*, 2013.

[28] G. Parseihian, S. Ystad, M. Aramaki, and R. Kronland-Martinet, "The process of sonification design for guidance tasks," *Wi: Journal of Mobile Media*, vol. 9, no. 2, 2015.

[29] L. J. Trainor, C. D. Tsang, and V. H. W. Cheung, "Preference for sensory consonance in 2- and 4-month-old infants," *Music Perception: An Interdisciplinary Journal*, vol. 20, no. 2, pp. 187–194, 2002.

[30] G. Dubus and R. Bresin, "A systematic review of mapping strategies for the sonification of physical quantities," *PLoS ONE*, vol. 8, no. 12, p. e82491, 2013.

[31] "Normal equal-loudness level contours – Acoustics International Organization for Standardization," 2003.

[32] S. S. Stevens, "The measurement of loudness," *Journal of the Acoustical Society of America*, vol. 27, no. 5, pp. 815–829, 1955.

[33] E. Terhardt, "Pitch of pure tones: its relation to intensity," in *Facts and models in hearing*. Springer, 1974, pp. 353–360.

[34] J. W. Beauchamp, "Synthesis by spectral amplitude and "brightness" matching of analyzed musical instrument tones," *Journal of the Audio Engineering Society*, vol. 30, no. 6, pp. 396–406, 1982.

[35] O. Lartillot, P. Toiviainen, and T. Eerola, "A matlab toolbox for music information retrieval," in *Data analysis, machine learning and applications*. Springer Berlin Heidelberg, 2008, pp. 261–268.

[36] R. W. Young, "Inharmonicity of plain wire piano strings," *Journal of the Acoustical Society of America*, vol. 24, no. 3, pp. 267–273, May 1952.

[37] H. Fletcher, E. D. Blackham, and R. Stratton, "Quality of piano tones," *Journal of the Acoustical Society of America*, vol. 34, no. 6, pp. 749–761, 1962.

[38] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[39] C. C. Wier, W. Jesteadt, and D. M. Green, "Frequency discrimination as a function of frequency and sensation level," *Journal of the Acoustical Society of America*, vol. 61, no. 1, pp. 178–184, 1977.

[40] J. A. Michon, "Studies on subjective duration: I. differential sensitivity in the perception of repeated temporal intervals," *Acta Psychologica*, vol. 22, no. 441–450, 1964.

[41] B. A. Wright, D. V. Buonomano, H. W. Mahncke, and M. M. Merzenich, "Learning and generalization of auditory temporal–interval discrimination in humans," *The Journal of Neuroscience*, vol. 17, no. 10, pp. 3956–3963, May 1997.

[42] S. Ferguson, D. Cabrera, K. Beilharz, and H.-J. Song, "Using psychoacoustical models for information sonification," in *Proceedings of the 12th International Conference on Auditory Display, London, UK*, 2006.

[43] M. A. Alonso-Arevalo, S. Shelley, D. Hermes, J. Hollowood, M. Pettitt, S. Sharples, and A. Kohlrausch, "Curve shape

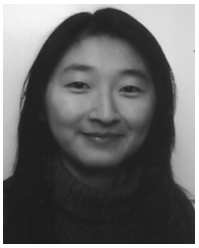
and curvature perception through interactive sonification," *ACM Transactions on Applied Perception*, vol. 9, no. 4, pp. 17:1–17:19, 2012. [Online]. Available: http://doi.acm.org/10.1145/2355598.2355600

[44] M. Aramaki, R. Kronland-Martinet, T. Voinier, and S. Ystad, "A percussive sound synthesizer based on physical and perceptual attributes," *Computer Music Journal*, vol. 30, no. 2, pp. 32–41, 2006.

[45] R. Kronland-Martinet, S. Ystad, and M. Aramaki, "High-level control of sound synthesis for sonification processes," *AI & society*, vol. 27, no. 2, pp. 245–255, 2012.

Gaëtan Parseihian received the Ph.D degree from the UPMC University, Paris, France, in 2012 for his work on binaural sonification for navigation aid. He is currently post-doc researcher at CNRS-LMA in the field of sound and computer science. His main research interests include sonification, auditory guidance, 3D sound, spatial perception, human computer interaction and augmented reality.

Charle Gondre received the Master degree in Art, Science, Technology from the Institute National Polytechnique de Grenoble in 2008. He has been working at the Laboratoire de Mcanique et d'Acoustique (Marseille, France) from 2008 to 2014. His interests include sound control and real-time synthesis algorithms. He currently works as freelance consultant and developer for the computer music industry.



Mitsuko Aramaki (M09-SM14) received the Ph.D. degree from the Aix-Marseille University, Marseille, France, in 2003, for her work on analysis and synthesis of impact sounds using physical and perceptual approaches. She is currently a researcher at the National Center for Scientific Research (CNRS). She has been working at the Institut de Neurosciences Cognitives de la Mditerrane (Marseille, France) from 2006 to 2011 and since 2012, she joined the Laboratoire de Mcanique et dAcoustique (Marseille, France). Her research mainly focused on sound modeling, perceptual and cognitive aspects of timbre, neuroscience methods and multimodal interactions in the context of virtual/augmented reality. She has been involved in industrial contracts with renowned companies such as Orange Labs and PSA Peugeot-Citroën. She is a member of the CMMR (Computer Music Multidisciplinary Research) steering committee and is regularly involved in the chair panel of the conference. She published 20 articles in international peer reviewed journals, more than 30 articles in international conference proceedings and co-edited 4 books published by Springer in their series LNCS.

Sølvi Ystad received her degree as a civil engineer in electronics from NTH (Norger Tekniske Hgskole), Trondheim, Norway in 1992 and her Ph.D. degree in acoustics from the Aix-Marseille University, Marseille, France, in 1998. She is currently a Researcher at the National French Research Center (CNRS) - Laboratory of Mechanics an Acoustics (LMA) in Marseille, France. Her research activities are related to sound modeling with a special emphasis on the identification of perceptually relevant sound structures to propose intuitive user interfaces for controlling synthesized sounds. She was in charge of the ANR funded research project Towards the sense of sounds, (http://www.sensons.cnrs-mrs.fr) from 2006-2009 and is currently participating in the ANR funded projects soniMove (http://sonimove.lma.cnrs-mrs.fr) and Potion (http://potion.cnrs-mrs.fr).

Richard Kronland-Martinet (M09-SM10) has a scientific background in theoretical physics, supplemented by a specialization in acoustics. In 1989 he obtained a Doctorate of Science (Habilitation Research) from the Aix-Marseille University, for his work on analysis, synthesis and processing of sounds by time-frequency and time-scale (wavelet) approaches. Since 1998 he is Director of Research at the Laboratoire de Mcanique et dAcoustique, CNRS, Marseille-France. His scientific activity concerns the science of sounds addressed from a multidisciplinary sight. He has been an important actor in the development of time-scale analysis methods and in their use for analysis, processing and synthesis of sound signals and music. The combination of physical concepts and methods of non-stationary signal processing has enabled him to develop many paradigms of synthesis, especially for instrumental sounds (piano, flute) but also for environmental sounds. His interest in perceptual and cognitive aspects associated with sounds have more recently led him to undertake research on the intuitive control of sounds over the process able to reproduce perceptual effects corresponding to high-level attributes.